

Multiscale Neural Nets for Image Processing

Tom Goldstein

...and...

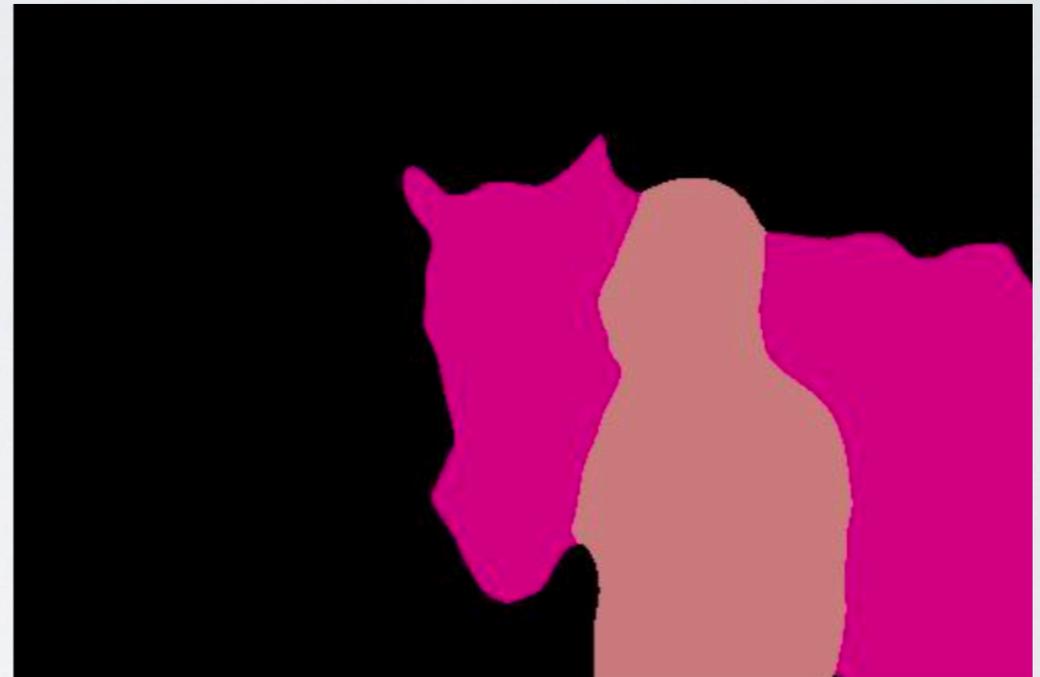
Sohil Shah, Pallabi Ghosh, Larry S. Davis



UNIVERSITY OF
MARYLAND

PROBLEM SETTING

PASCAL Visual Object Classes (VOC)



Cityscapes



WHAT MAKES SEGMENTATION HARD?

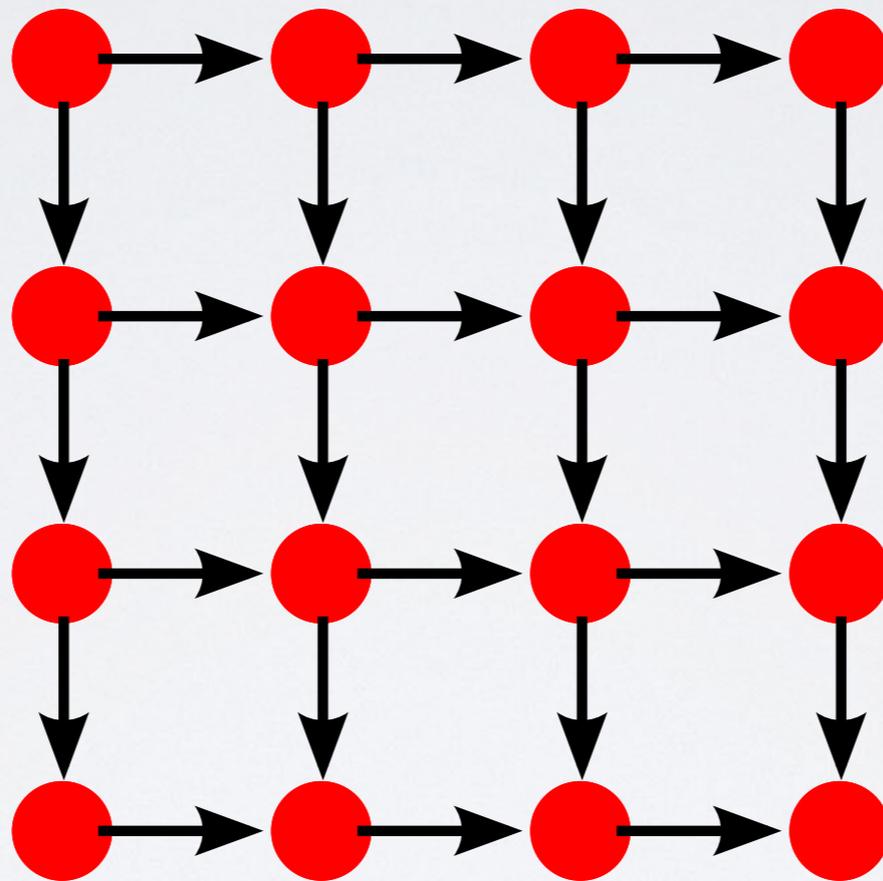
Field of view problem

High-resolution outputs

Small data

TV IN 2D

$$(\nabla x)_{ij} = (x_{i+1,j} - x_{i,j}, x_{i,j+1} - x_{i,j})$$



Anisotropic $|(\nabla x)_{ij}| = |x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}|$

Isotropic $\|(\nabla x)_{ij}\| = \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}$

CONVEX SEGMENTATION: FINE SCALE

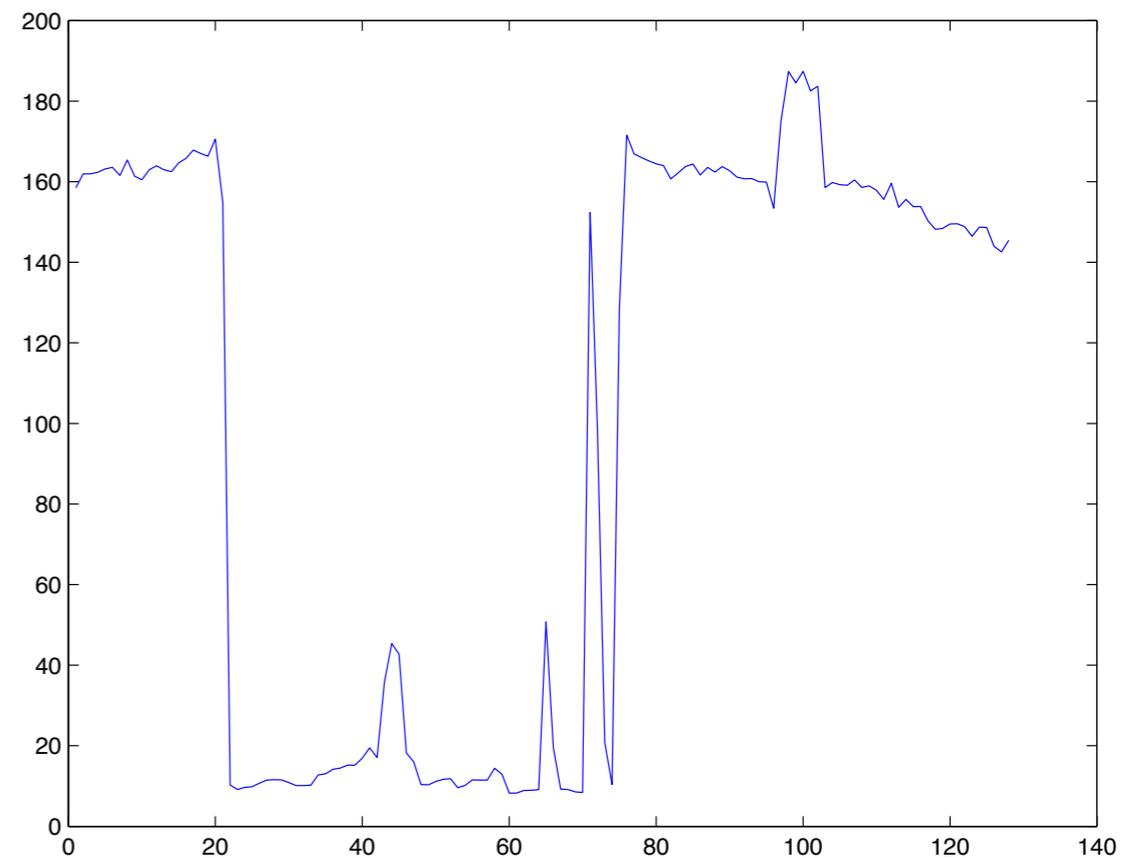
- Solve speed limited by the **CFL**

$$\min \mu |\nabla u| + \|u - f\|^2$$



This

Depends on This



CONVEX SEGMENTATION: COARSE SCALE

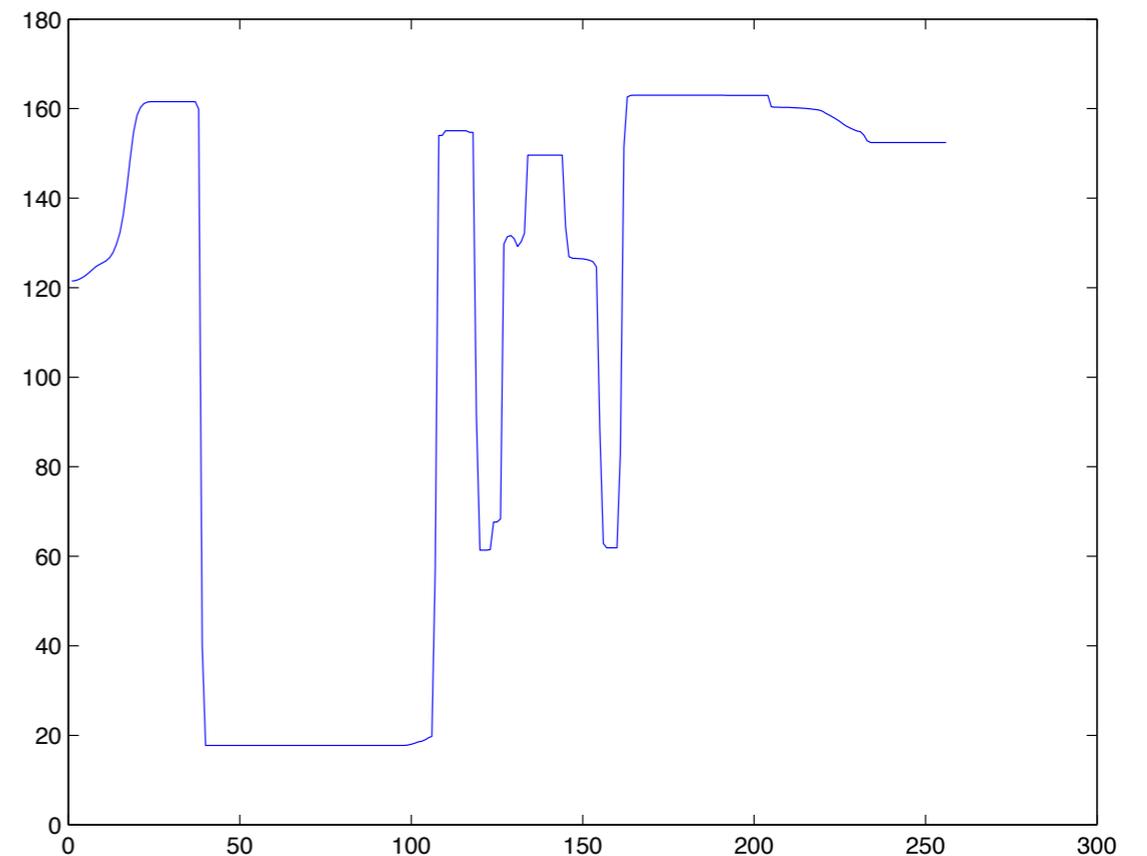
- TV changes the CFL condition

$$\min \mu |\nabla u| + \|u - f\|^2$$



This

Depends on This



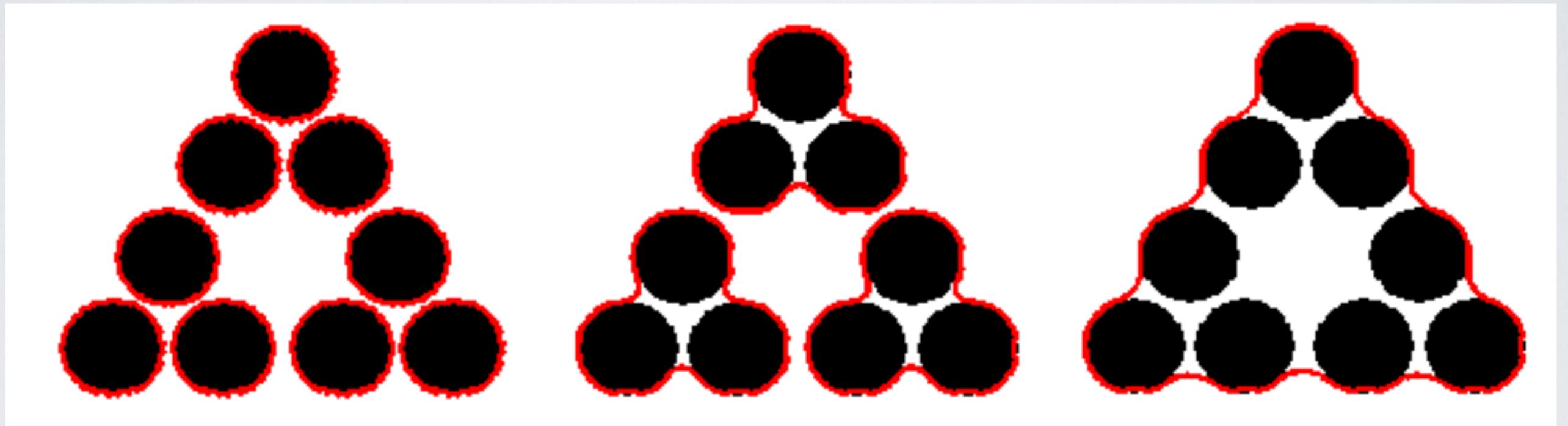
SCALE AFFECTS SPEED!

Segmentation using TV

15 iterations

280 iterations

4500 iterations



Solution: use multi-grid!

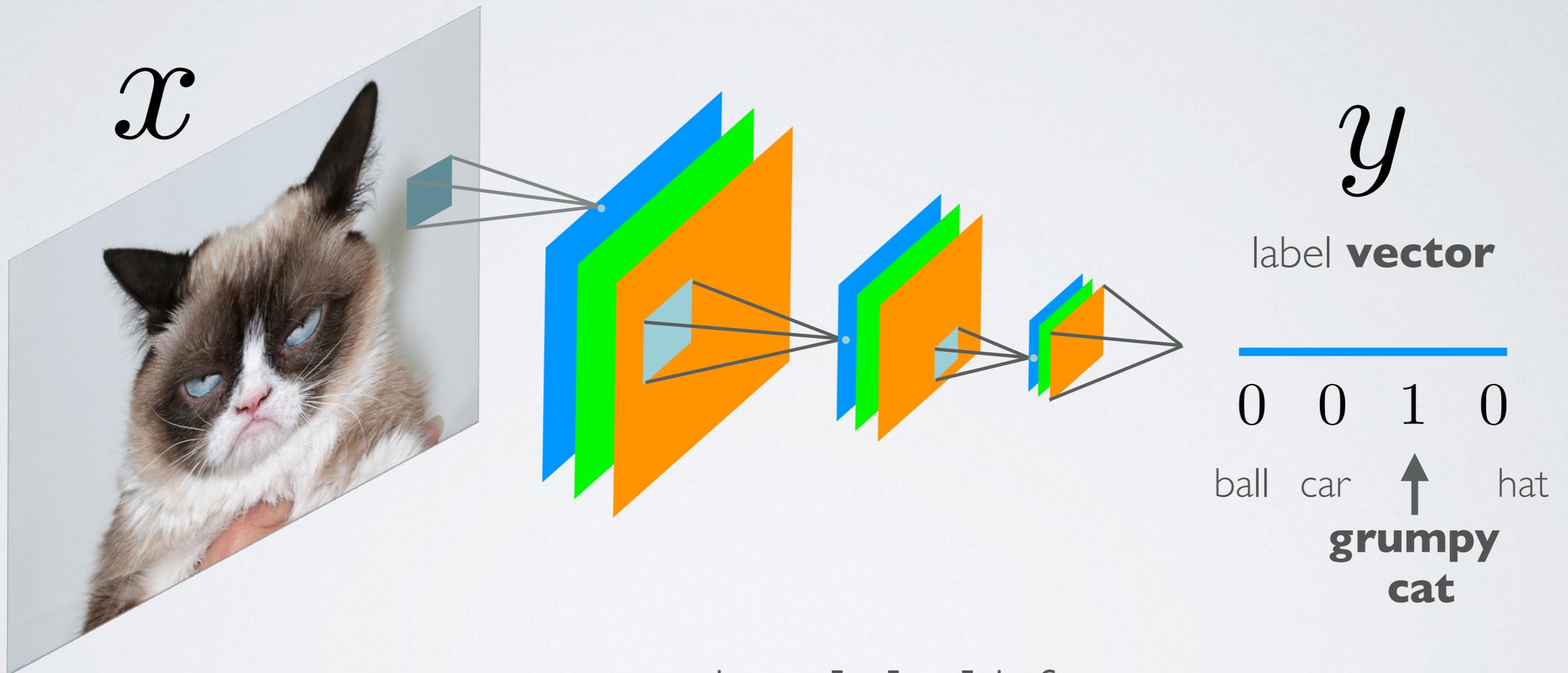
NEURAL NETS HAVE THE SAME PROBLEMS



far away points
talk to each other

CLASSIFICATION NET

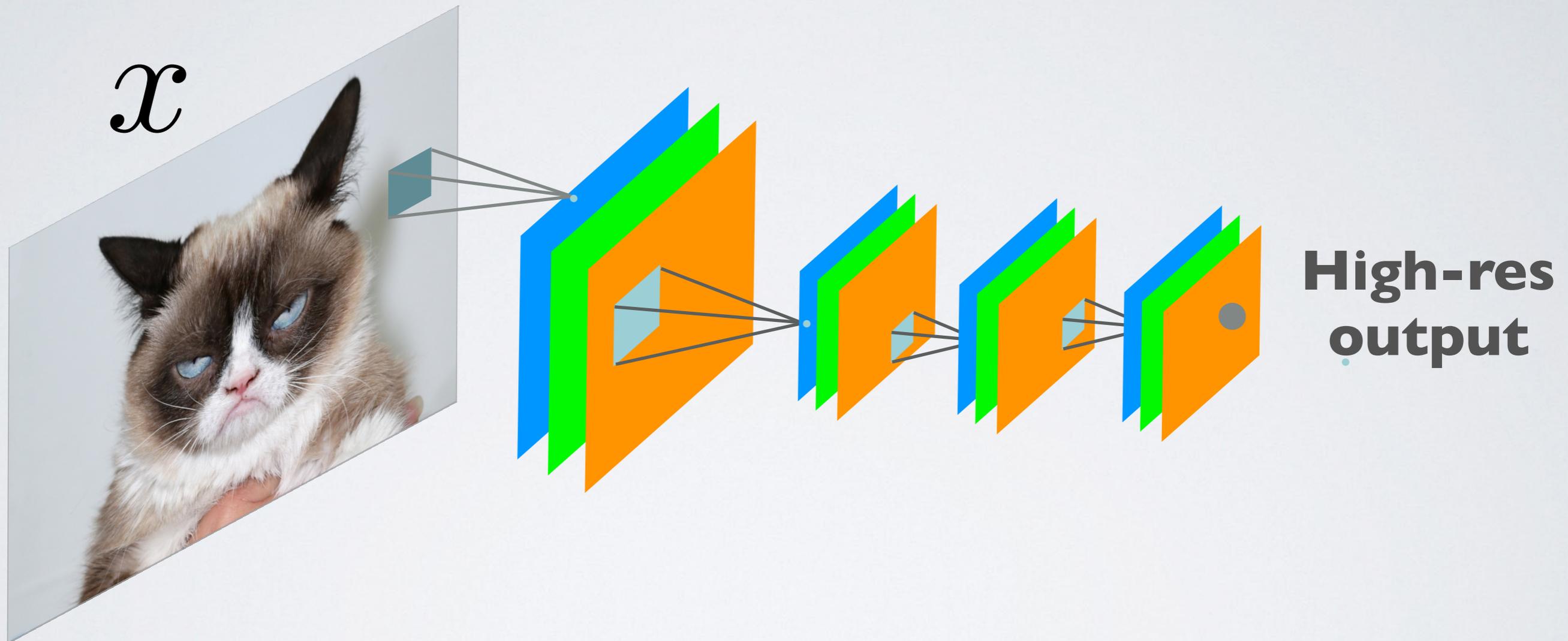
Globalize via **pooling**



output contains **global** info

CLASSIFICATION NET

Globalize via **convolution**

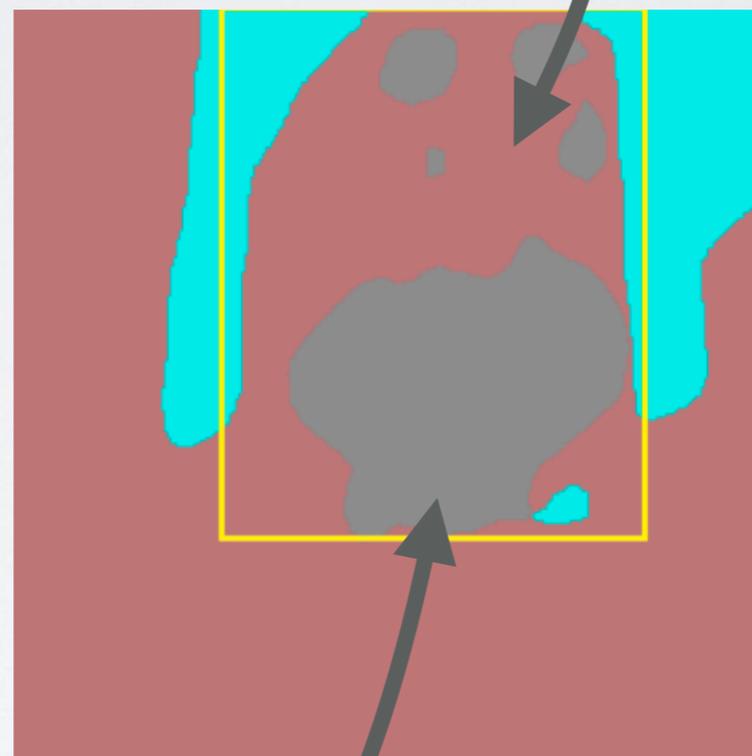


CFL condition becomes **field of view problem**

FIELD OF VIEW BREAKDOWN



Building



Skyscraper

sky
tree
grass
earth
plant
car
boat
water
river
house
building
skyscraper
wall
floor
bed
cabinet
table
curtain
lamp
pillow

Example from: **Zhao et al. "Pyramid Scene Parsing Network"**

GLOBALIZATION METHODS

THINGS GET COMPLICATED...

APPROACHES TO GLOBALIZATION

Dilated/de-conv modules

Chen et al. “Semantic image segmentation...” 2014

Yu & Koltun, “Multi-scale context aggregation” 2015

Multi-scale feature ensembling

Long et al., “Fully Convolution Semantic Segmentation” 2015

Chen et al. “Attention to scale...” 2016

Xia et al. “Zoom better to see clearer” 2016

Hariharan et al. “Hypercolumns for object segmentation” 2015

Hand-crafted features

Lazebnik, Schmid, Ponce, “Beyond bags of features” 2006

Lucchi et al. “Are spatial constraints necessary for segmentation” 2011

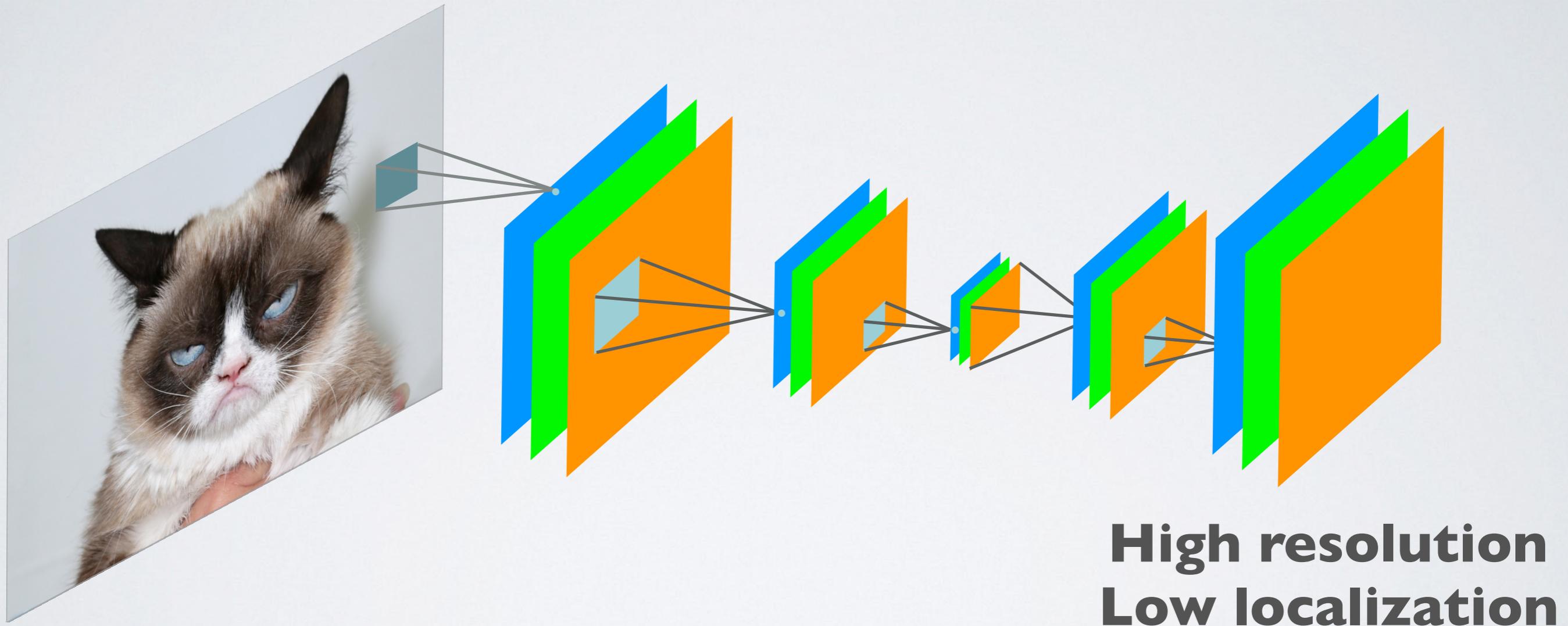
Chen et al. “DeepLab”

Hengshuang et al. “PSPNet”

SEGMENTATION NET

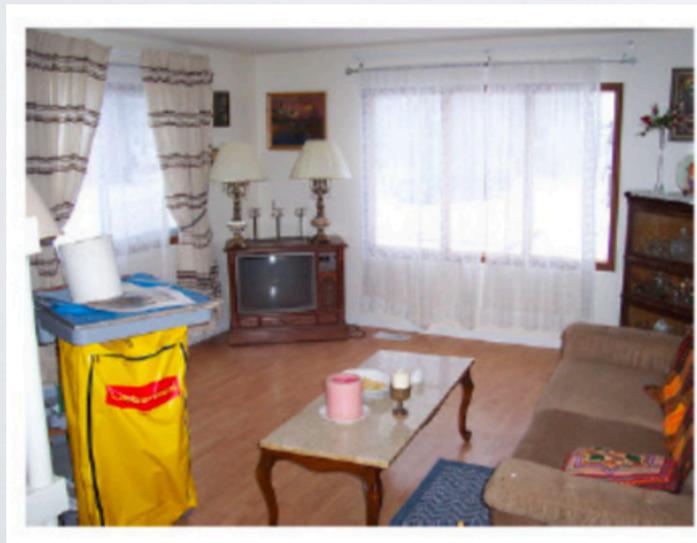
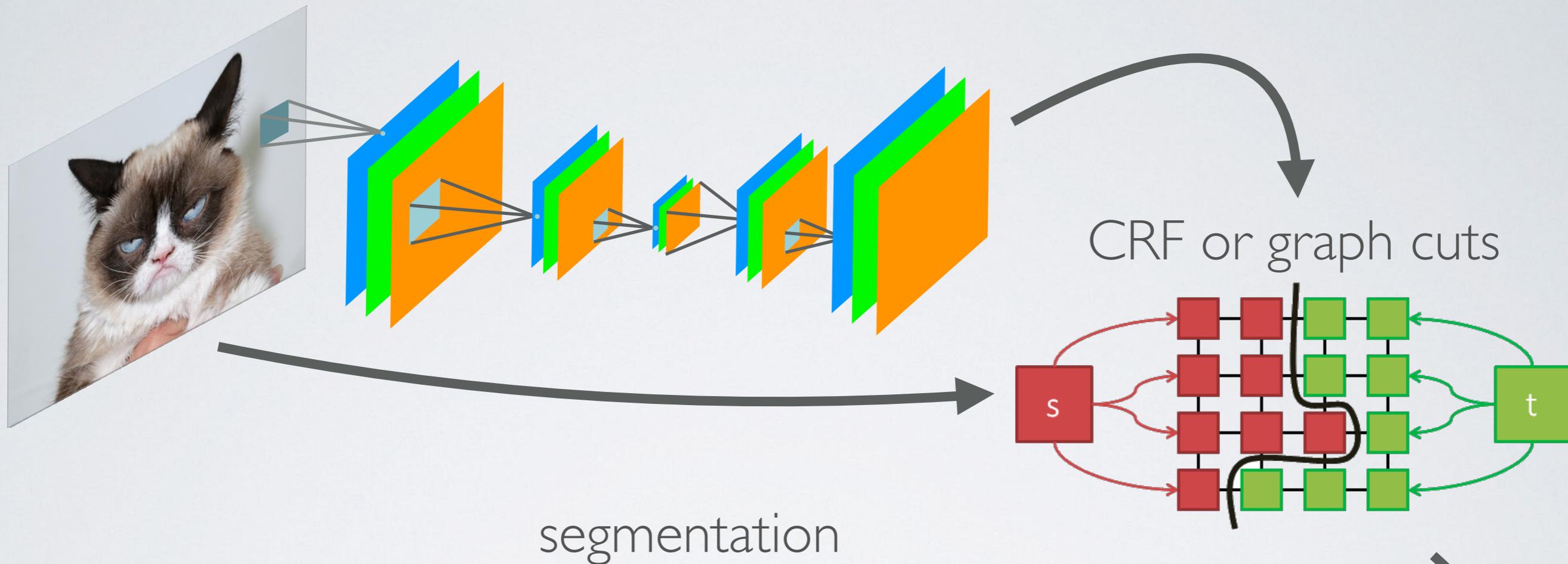
standard net

de-conv units



LC Chen et al. "DeepLab"

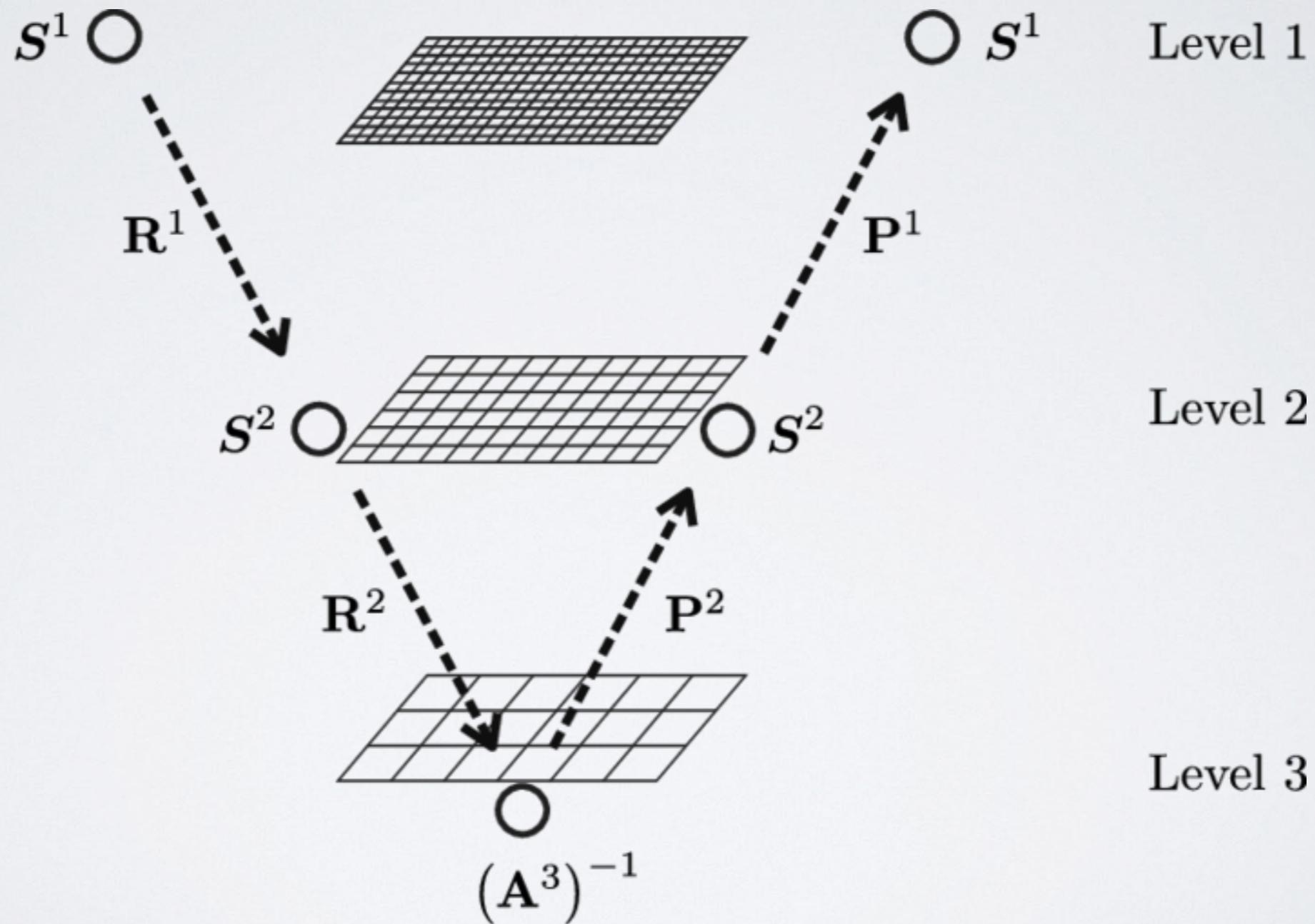
DEEPLAB



BACK TO BASICS...

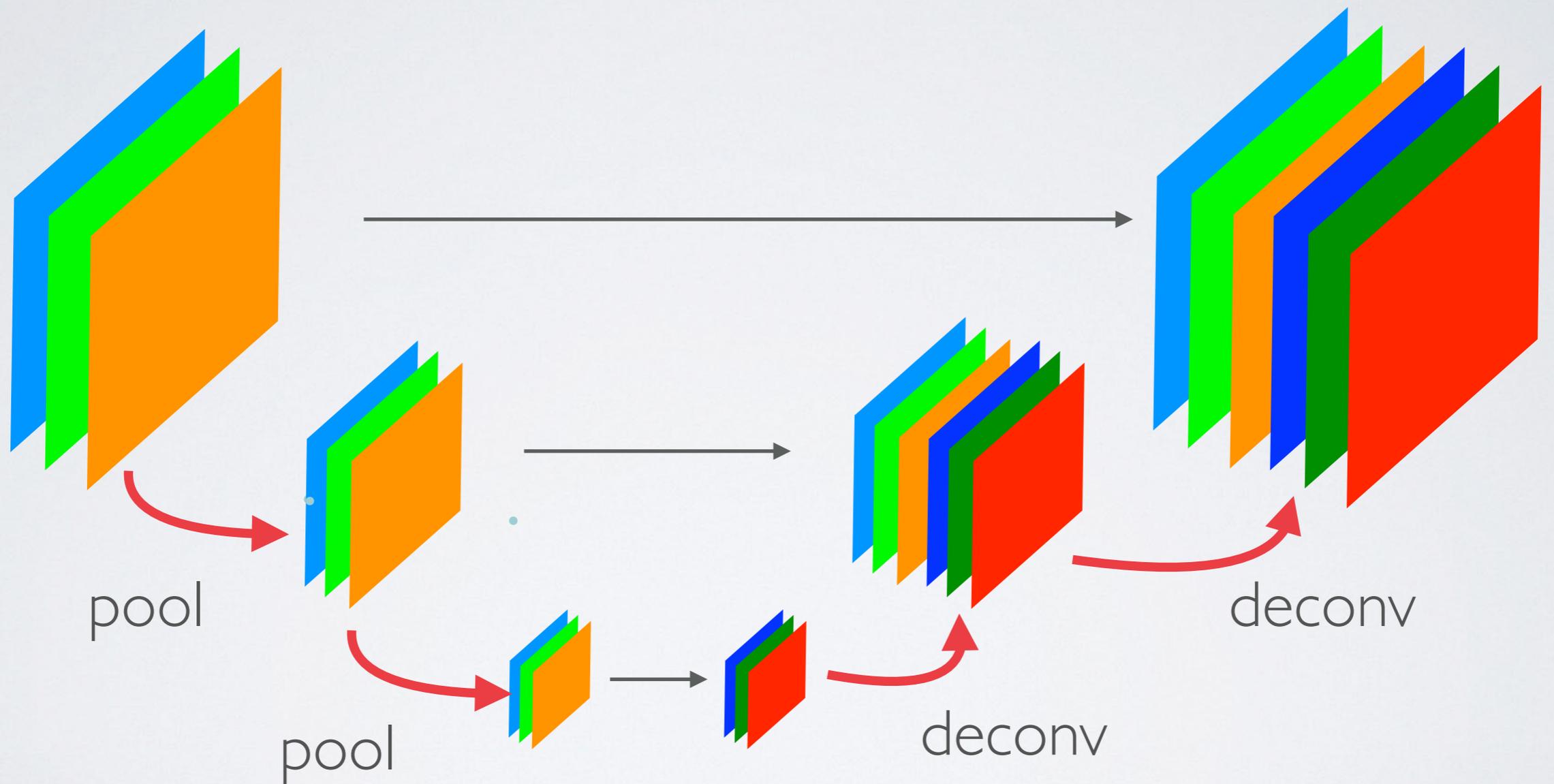
A “NO-FRILLS” APPROACH

MULTI-GRID SOLVERS



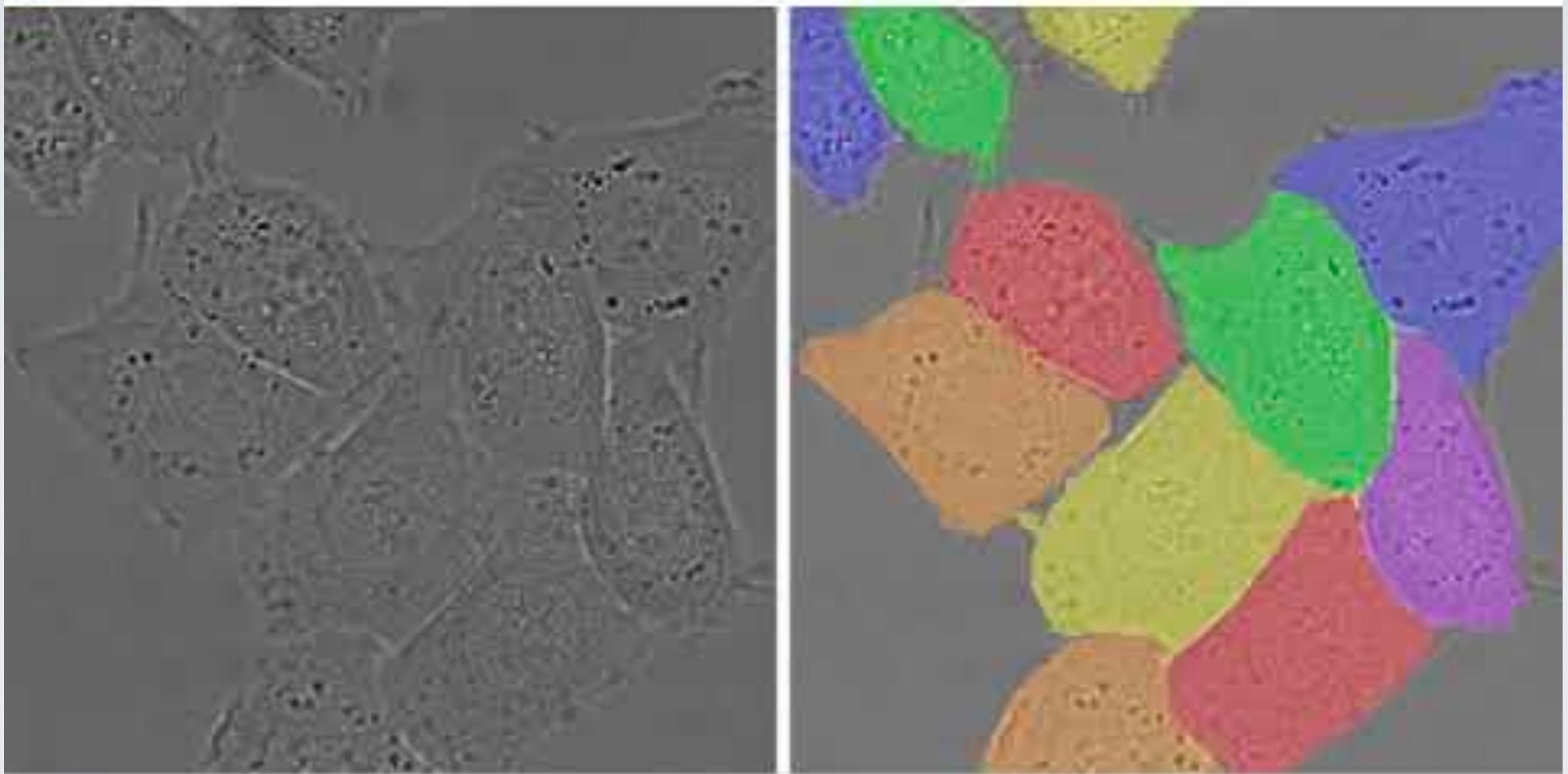
THE U-NET

A V-cycle for neural nets



MEDICAL IMAGING

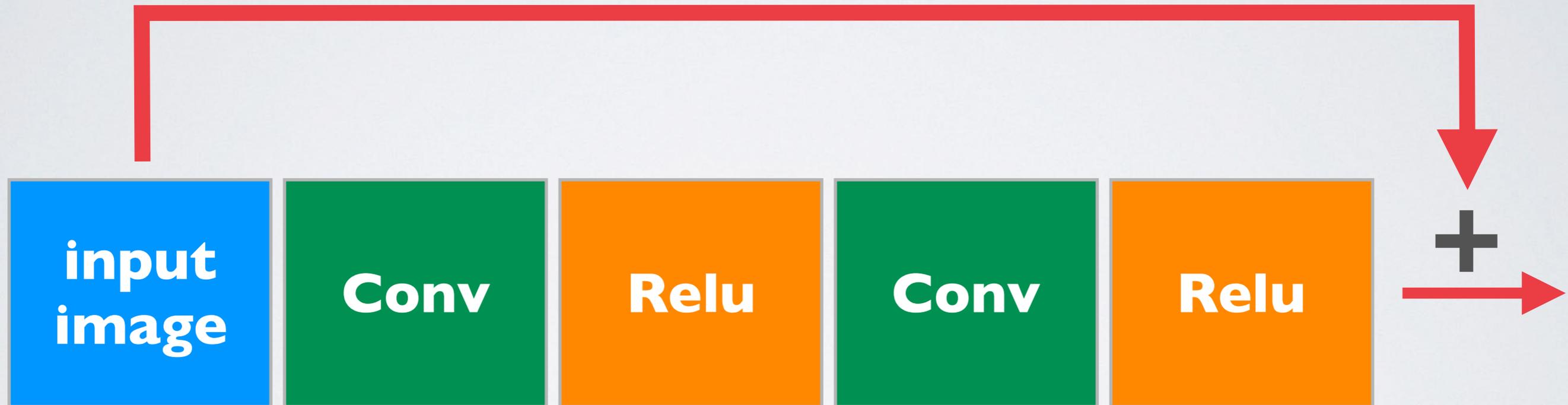
resolution is **high**
class complexity is **low**



NEURAL NET BUILDING BLOCKS

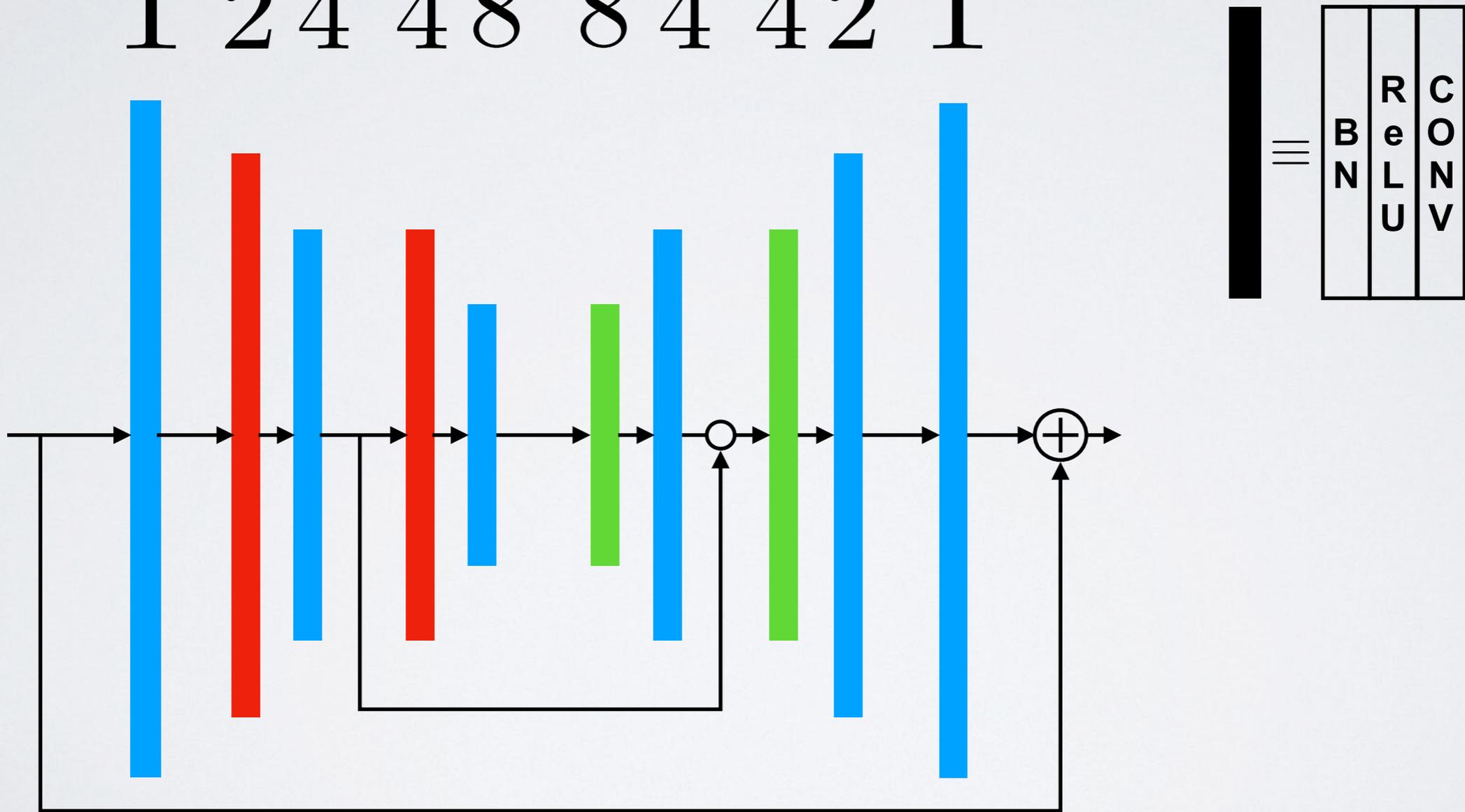


“RESNET” BLOCK



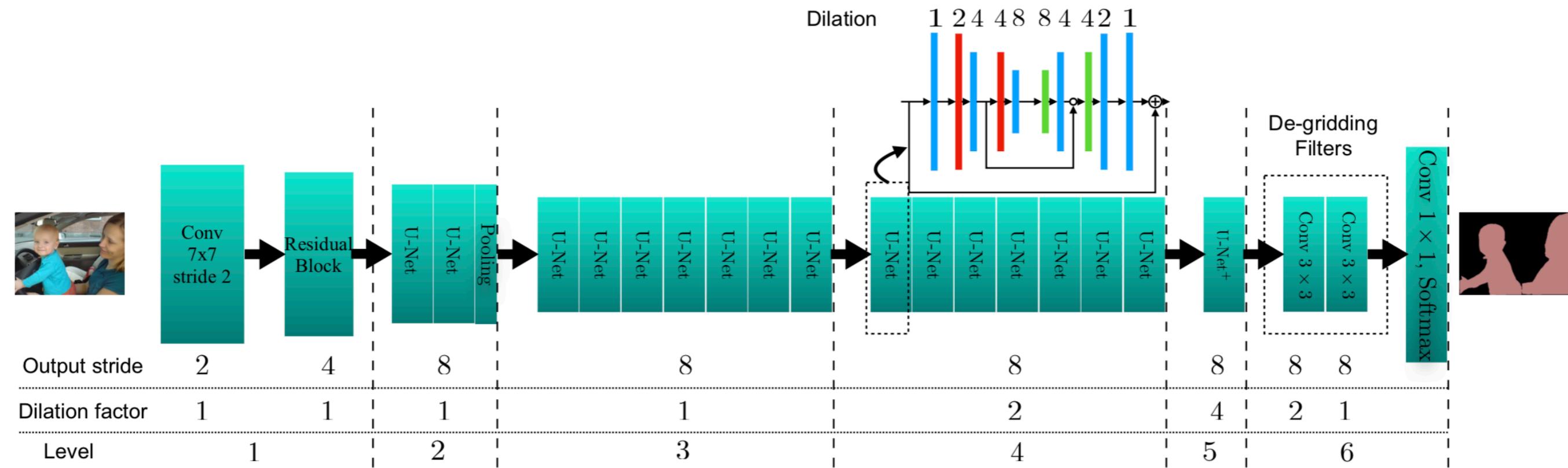
MODIFIED U-NET BLOCK

1 2 4 4 8 8 4 4 2 1



BIG IDEA: STACKED U-NETS

Combine information globalization of U-nets with power of ResNets



No frills!

TRAINING ON REAL DATASETS

LIMITED TRAINING DATA

ImageNet 1,000,000

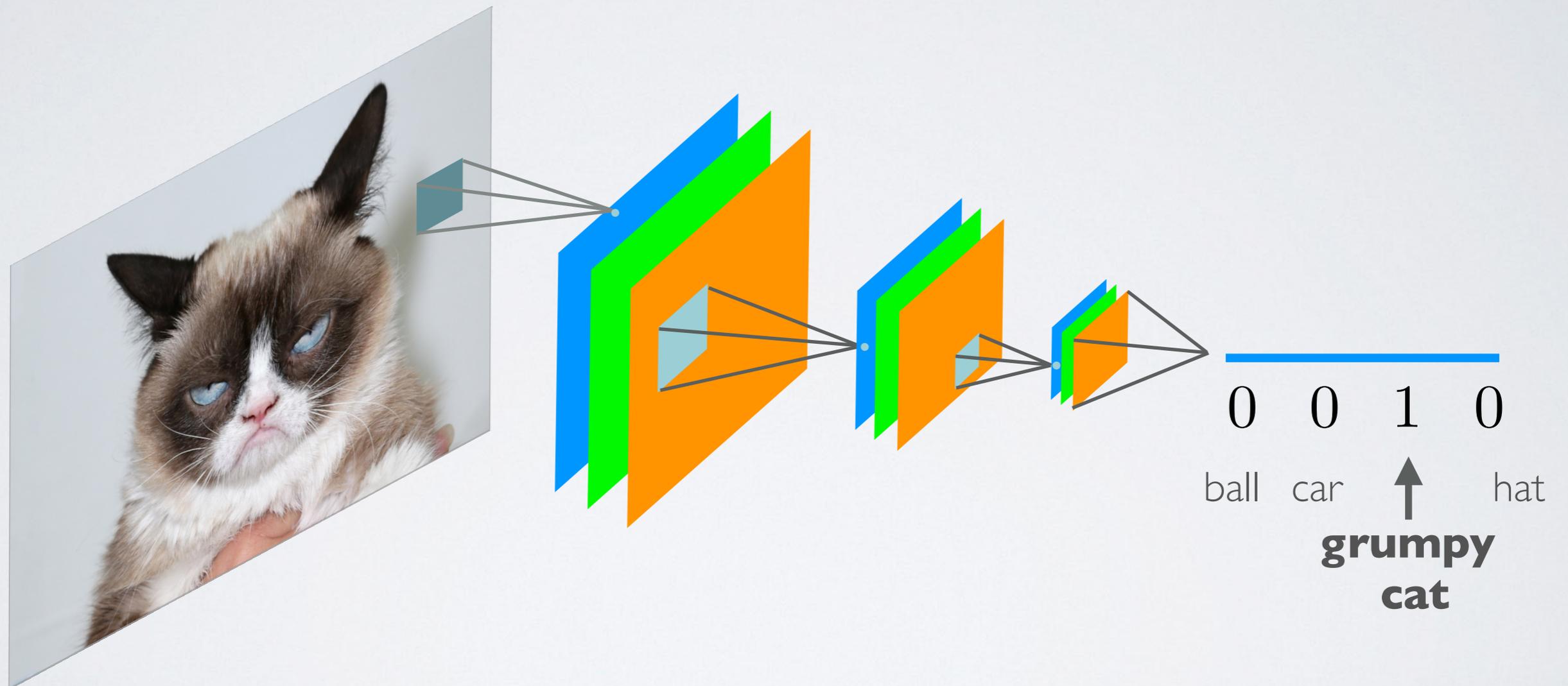
MS-COCO 90,000

PASCAL VOC 1464

Solution: pre-train on big datasets, fine tune on small

STANDARD APPROACH

Phase I: Train a classification net

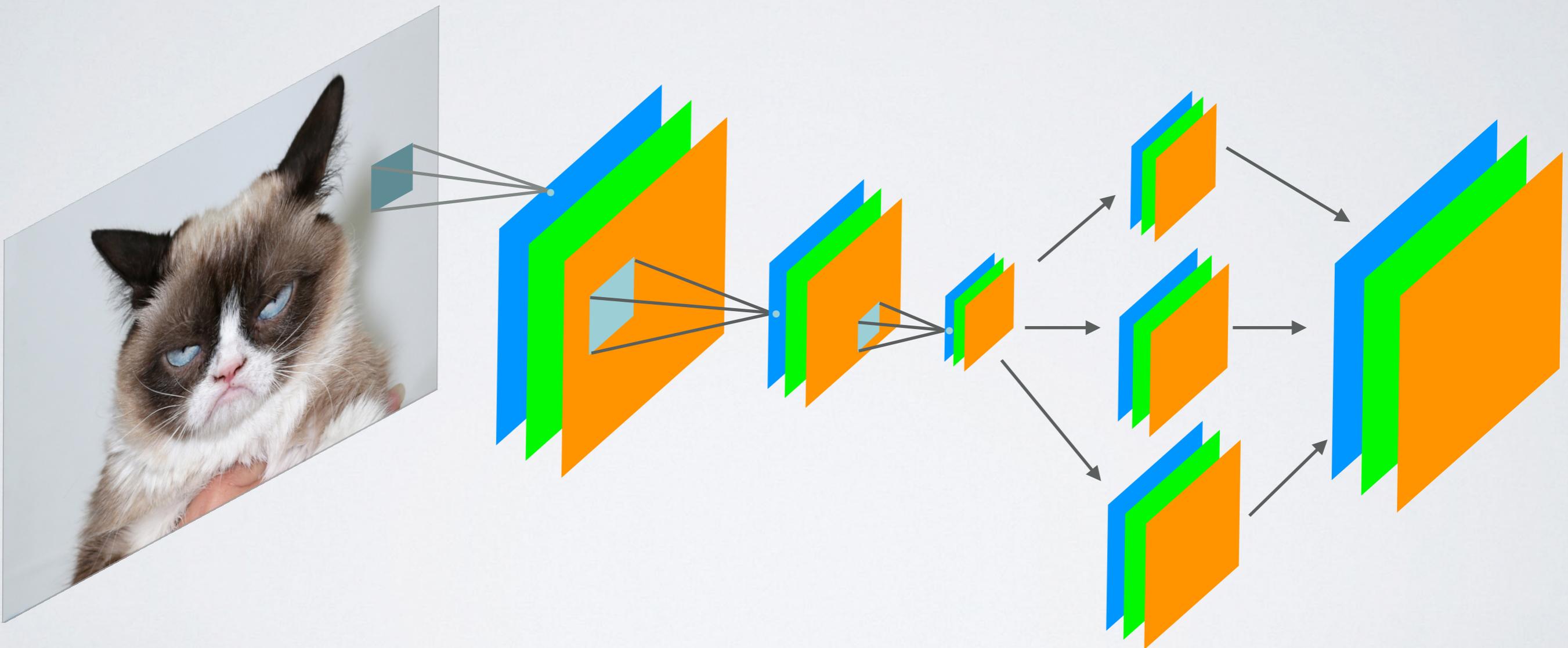


STANDARD APPROACH

Phase 2: Add fancy stuff

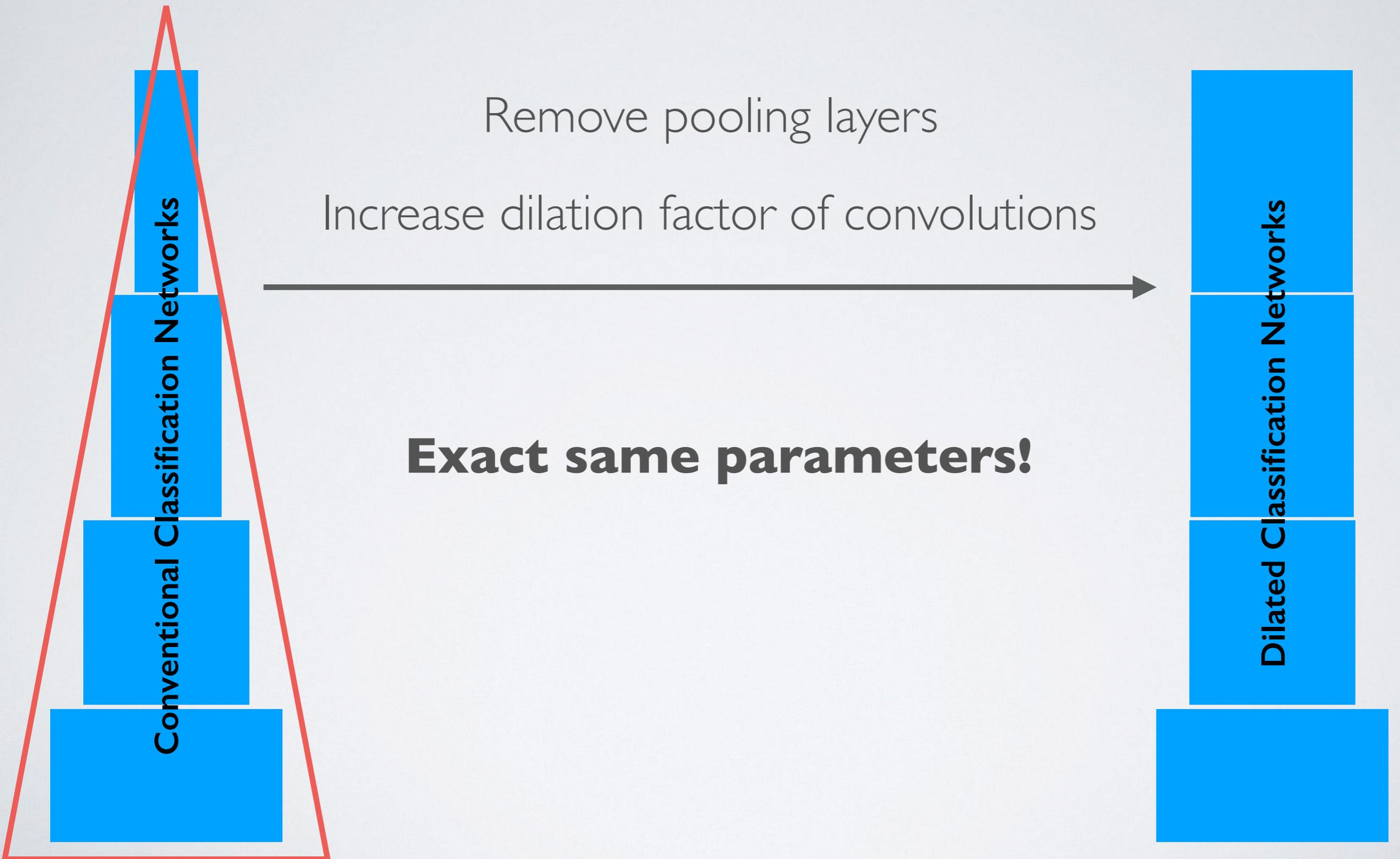
standard net

fancy stuff



Roughly doubles parameters!

Dilated SUNets



RESULTS

METRICS

Classification

Top-1 accuracy

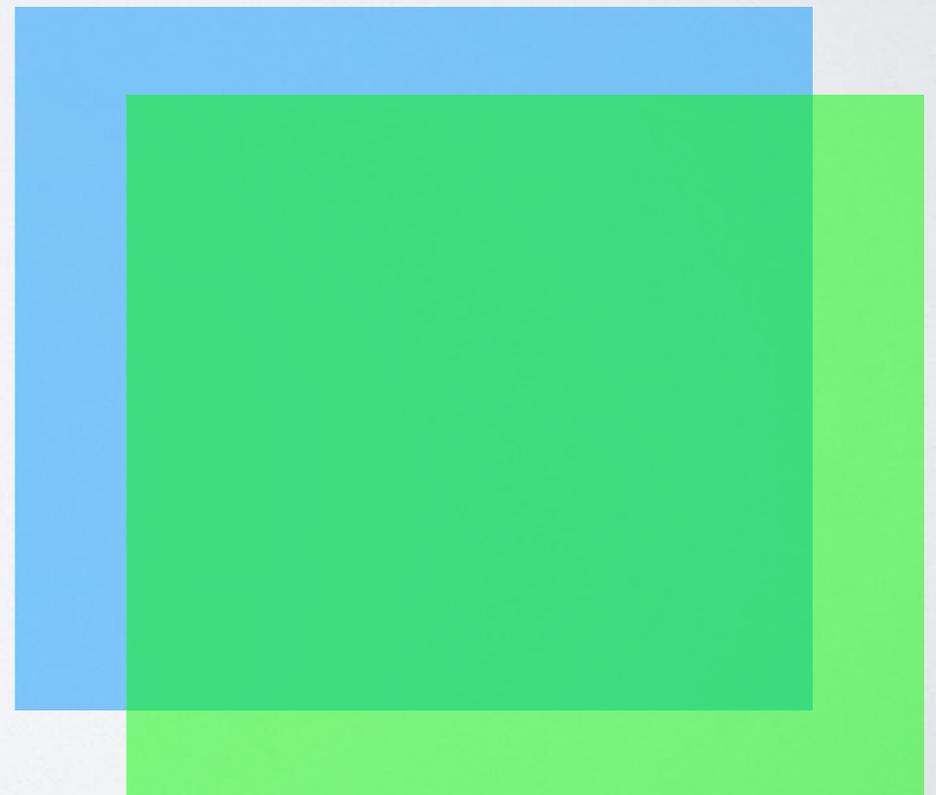
Top-5 accuracy

“Cat”



Segmentation

Intersection over union



RESULTS

Classification Performance

Model	Top-1	Top-5	Depth	Params
ResNet-18 [†]	30.24	10.92	18	11.7M
ResNet-50 [†]	23.85	7.13	50	25.6M
ResNet-101 [†]	22.63	6.44	101	44.5M
DenseNet-201 [†]	22.80	6.43	201	20M
DenseNet-161 [†]	22.35	6.20	161	28.5M
SUNet-64	29.28	10.21	111	6.9M
SUNet-128	23.64	7.56	111	24.6M
SUNet-7-128	22.47	6.85	171	37.7M

Segmentation Performance

Model	mIoU
ResNet-101 [218]	68.39
SUNet-64	72.85
SUNet-128	77.16
SUNet-7-128	78.95

- ✓ 4.5% mIoU ↑ with **7x** fewer parameters
- ✓ 10.5% mIoU ↑ w/o sacrificing classification performance

Sohil Shah, Pallabi Ghosh, Larry S. Davis and Tom Goldstein, "Stacked U-Nets: A No-Frills Approach to Natural Image Segmentation"

SEGMENTATION RESULTS

Training on train-set

Methods	Validation mIoU
Resnet+ASPP	82.70
Xception+ASPP+Decoder	83.34
SUNet-7128	83.27

Training on train+val (2x data)

Methods	mIoU
Piecewise (VGG16) [224]	78.0
LRR+CRF [209]	77.3
DeepLabv2+CRF [210]	79.7
Large-Kernel+CRF [220]	82.2
Deep Layer Cascade* [254]	82.7
Understanding Conv [221]	83.1
RefineNet [205]	82.4
RefineNet-ResNet152 [205]	83.4
PSPNet [217]	85.4
SUNet-7-128	84.3

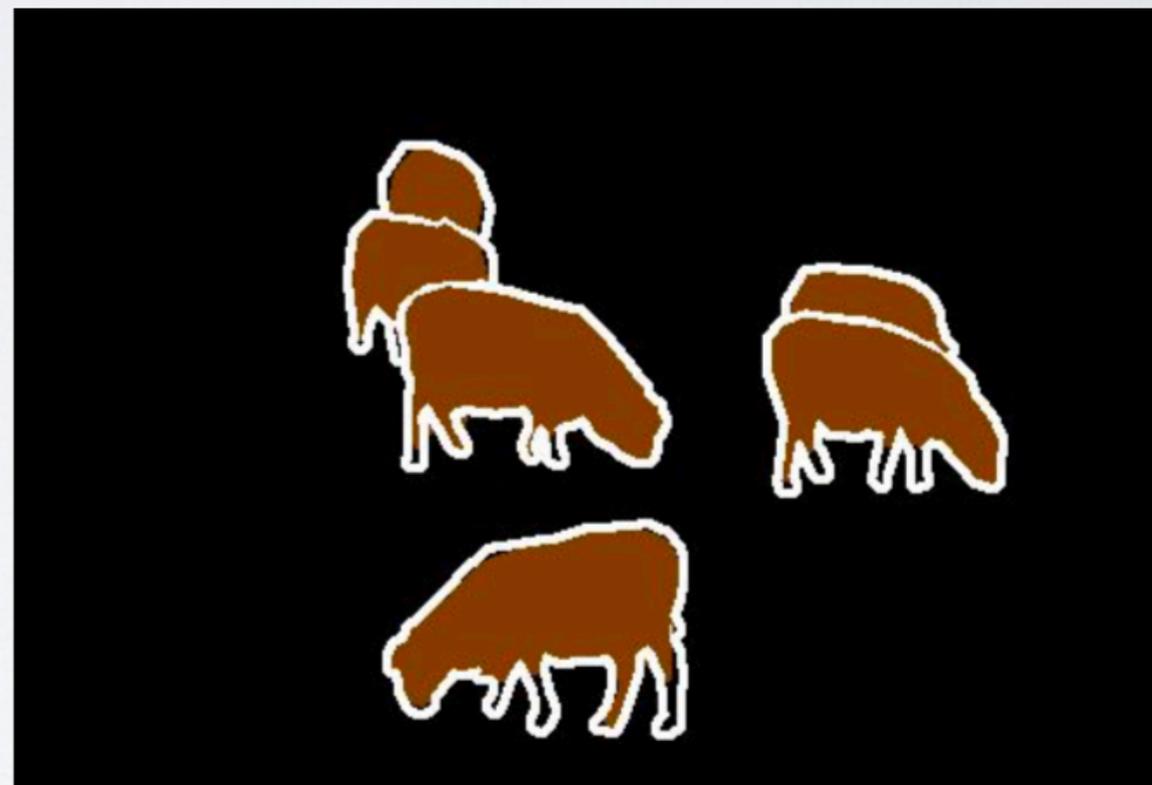
Advantages...

1/2 the weights, 1/5 the RAM

SAMPLE RESULTS



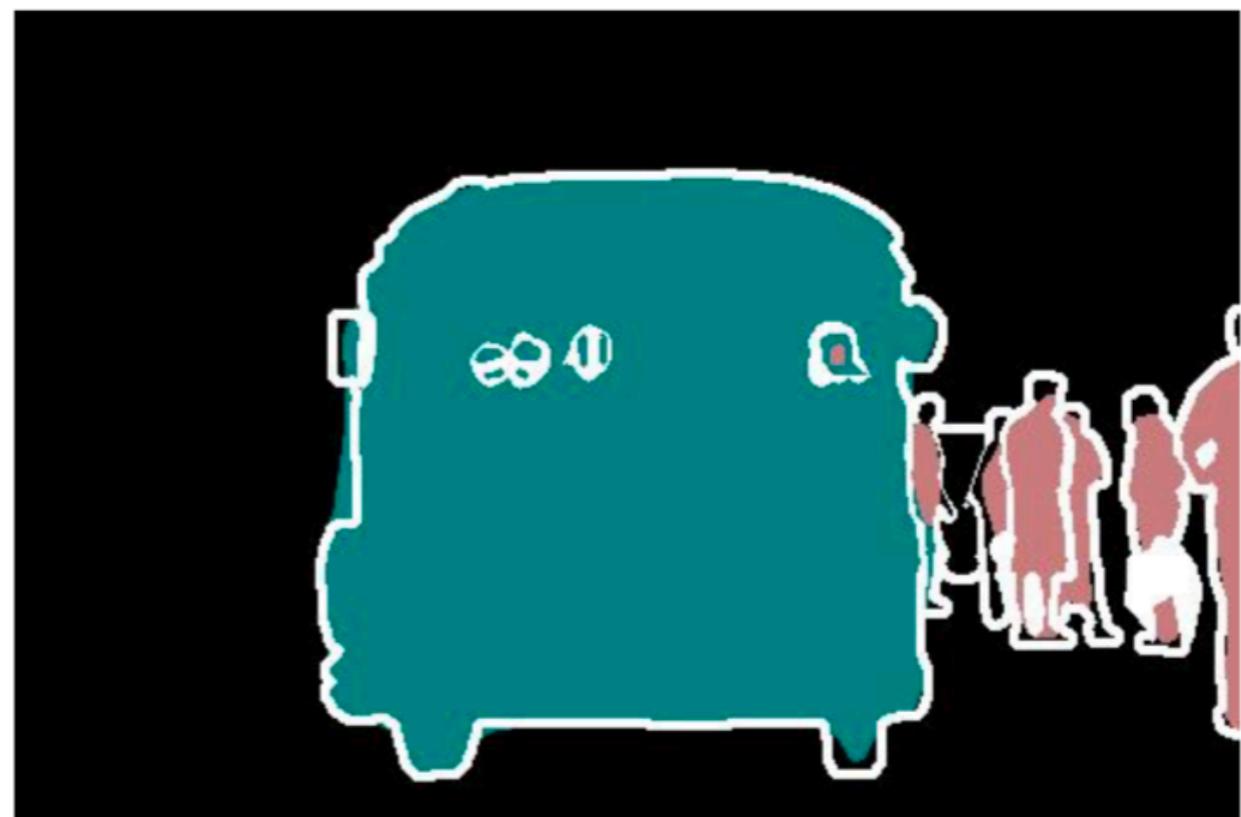
SAMPLE RESULTS



SAMPLE RESULTS



SAMPLE RESULTS



FAILURE CASE

input



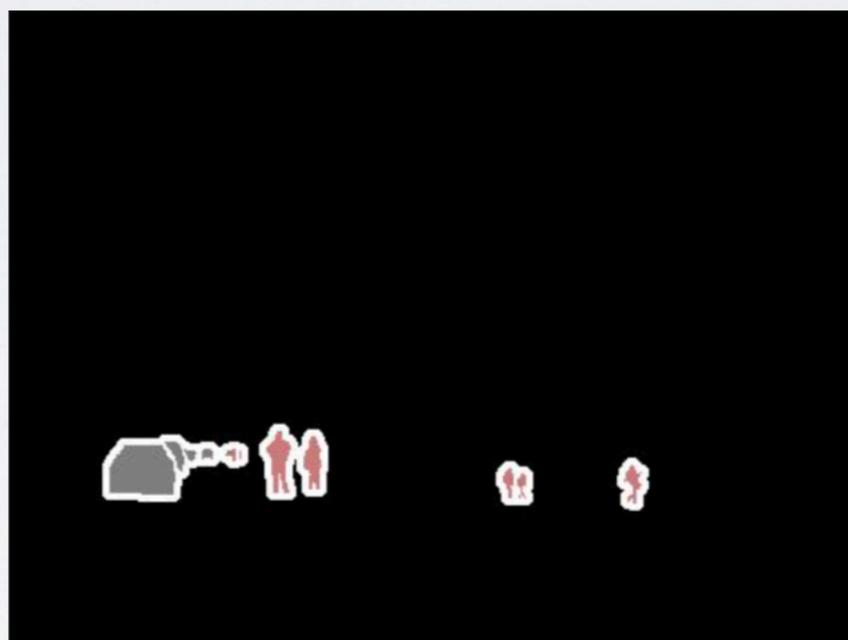
target



output



FAILURE CASE



THANKS!

Stacked U-Nets: A No-Frills Approach to Natural Image Segmentation

Sohil Shah, Pallabi Ghosh, Larry S. Davis and Tom Goldstein

